CCT College Dublin Continuous Assessment

Programme Title:	HDip in Science in DAB/ AI Applications		
Delivery Mode:	FT/ PT		
Cohort Details:	*HDip in Data Analytics & HDip AI Apps- Sept 2024 - HCI cohort, Semester 1*		
Module Title(s):	*Machine Learning*		
Assignment Type:	*Individual*	**Weighting(s):**	*50%*
Assignment Title:	*CA1*		
Lecturer(s):	*Dr Muhammad Iqbal*		
Issue Date:	*1st October 2024*		
Submission Deadline Date:	*17th November 2024 (23:59 IST)*		
Late Submission Penalty:	Late submissions will be accepted up to **5** calendar days after the deadline. All late submissions are subject to a penalty of **10%** of the mark awarded. Submissions received more than 5 calendar days after the deadline above **will not** be accepted and a mark of 0% will be awarded.		
Method of Submission:	**This assignment is submitted via Moodle.**		
Instructions for Submission:	*Upload all files separately, such as MS word file, jupyter notebook, dataset/datasets and any supporting information on Moodle.*		
Feedback Method:	**Results posted in Moodle gradebook**		
Feedback Date:	*8th December 2024*		

Assessment Outline

This is an individual CA1 assignment for machine learning using the PYTHON programming language. Develop and deploy machine learning models in any one of the following areas only, analyse and subsequently interpret the results.

- Education
- Justice, Legal system, and Public Safety
- Housing and Zoning

You can find any public dataset from an authentic resource repository and the dataset should have at least 5 columns after cleaning and more than 300 rows.
The type of question(s) that you should formulate for the assignment will depend on the chosen area of the dataset that you are considering for the machine learning assignment.
Suggested possible analysis / assignment questions are mentioned below (this is a small, suggested, sample of questions, other questions may be more appropriate to your assignment)

- What are the most important features for predicting X as a target variable?
- Which classification approach do you prefer for the prediction of X as a target variable, and why?
- How to classify the loyal and churn customers using Support Vector Machines?
- Why is dimensionality reduction important in machine learning?

The student would need to consider the following instructions (a - d) during the development of this assignment.

a) Logical justification based on the reasoning for the specific choice of machine learning approaches.
b) Multiple machine learning models (at least two) using hyperparameters and a comparison between the chosen modelling approaches.
c) Visualise your comparison of ML modelling outcomes. You may use a statistical approach to argue that one feature is more important than other features.
d) Cross-validation methods should be used to justify the authenticity of your ML results.

You will present the findings and defend the results in the report (MS Doc) by highlighting your work. Your report should capture the following aspects that are relevant to your assignment investigations.

1) A clear introduction, motivation, a description of the problem domain, and an explanation of how the assignment's goals are justified using Prediction / Classification algorithms.

(20 marks)

2) Characterization of data, pre-processing, explanation and description of techniques used for the variation in the accuracy across three training splits (10%, 15% and 25%) using cross validation techniques.

(30 marks)

3) What is the primary purpose of hyperparameter tuning in machine learning? Could you elaborate on specific hyperparameter tuning techniques (e.g., GridSearchCV) applied to machine learning models to find optimal parameters?

(25 marks)

4) Interpret and explain the results obtained, discuss overfitting / underfitting / generalisation, provide a rationale for the chosen models and use visualisations to support your findings. Comments in Python code, conclusions of the assignment should be specified at the end of the report. Harvard Style must be used for citations and references.

(25 marks)

Assessment Requirements

All assessment submissions must meet the following minimum requirements:

- Be submitted in the format outlined in the assignment summary table.
- The code and datasets should be provided and uploaded on Moodle.
- Maximum number of in the report should be in between 1000 to 1050 excluding the citations and references.
- Be submitted by the deadline date specified or be subject to late submission penalties.
- Be submitted via Moodle upload.
- Use Harvard Referencing when citing third party material.
- Be the student's own work.
- Include the CCT assessment cover page.
- Students must use the classroom GitHub link (https://classroom.github.com/a/RUOsR2q-) and you should have more than 7 commits on GitHub on different dates and times.
- Use of Gen AI (chatGPT or Copilot or others) must be clearly stated for any part of your assignment.

Learning Outcomes:

- This assessment addresses the following module learning outcomes for this module

- Develop a machine learning strategy for a given domain, communicate this strategy effectively to team members, peers and project stakeholders (CRISP-DM) (Linked to PLO 1, PLO 4, PLO 6)
- Implement a range of classification and regression techniques and detail /document their suitability for a variety of problem domains. (Linked to PLO 5)
- Critically evaluate and optimise the performance of Machine Learning models. (Linked to PLO 3)

Statement of Acceptable Use of Artificial Intelligence

Acceptable and Unacceptable Use of AI

- The use of generative AI tools (e.g. ChatGPT, Dall-e, etc.) is permitted in this assignment for the following activities:
  - Brainstorming and refining your ideas;
  - Fine tuning your research questions;
  - Finding information on your topic;
  - Drafting an outline to organise your thoughts; and
  - Checking grammar and style.
- The use of generative AI tools is not permitted in this course for the following activities:
  - Impersonating you in classroom context
  - Completing group work that your group has assigned to you
  - Generating code for your assignment
  - Writing a draft of a writing assignment
  - Writing entire sentences, paragraphs or papers to complete class assignments.
- You are responsible for the information you submit based on an AI query. Your use of AI tools must be properly documented and cited.
- Any assignment that is found to have used generative AI tools in an unauthorised way will be subject to college disciplinary procedures as outlined in the QA Manual.
- When in doubt about permitted usage, please ask for clarification.

Grading Criteria

This grading rubric sets out the marking criteria for your assignment.

Criteria	Criteria 1	Criteria 2	Criteria 3	Criteria 4
Marking Criteria	**Introduction to Problem Description, Motivation, and Objectives**	**Characterization and Cleaning of Dataset, Training and Testing of ML Models**	**Purpose of Hyperparameter Tuning and Application of Hyperparameter Tuning Technique**	**Interpretation of Results, Code Description, Comments, Conclusions, Citations, and References**
Weighting per Criteria	**20%**	**30%**	**25%**	**25%**
Excellent (+70%)	An excellent introduction to problem description and motivation that provide a precise and clear case for the proposed Machine Learning assignment. An excellent specification of objectives succinctly.	An excellent characterization and cleaning of the dataset that summarizes all details from source to fields. An excellent accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the generalizability of the model and it should justify the results in an excellent way.	Describes the purpose of hyperparameter tuning with clarity and describes hyperparameter tuning techniques with clarity, though may have some minor gaps or less detailed explanation.	An excellent interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An excellent defence is provided.
Very Good (60-69%)	A very good introduction to problem description and motivation that provides a very convincing case for the proposed Machine Learning assignment. A very good specification of objectives.	A very good characterization and cleaning of the dataset that summarizes all details from source to fields. A very good accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the partial generalizability of the model and it should justify the results.	Conveys a basic understanding of the purpose of hyperparameter tuning but lacks depth or may contain inaccuracies. Conveys a basic understanding of hyperparameter tuning techniques but lacks depth or may contain inaccuracies.	A very good interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results state that the models are neither overfitted nor underfitted. A very good justification is provided.

Good (50-59%)	A good introduction to problem description and motivation that furnishes a largely convincing case for the proposed Machine Learning assignment. A good specification of objectives.	A good characterization and cleaning of the dataset that summarizes all details from source to fields. A good accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the partial generalizability of the model.	Mentions the purpose of hyperparameter tuning but with significant gaps or inaccuracies, demonstrating a limited understanding. Mentions hyperparameter tuning techniques but with significant gaps or inaccuracies, demonstrating a limited understanding.	A good interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results state that the models are overfitted but not under fitted. A good justification is provided.
Acceptable (40-49%)	An acceptable introduction to problem description and motivation that offers a somewhat weak case for the proposed Machine Learning assignment. An adequate specification of objectives.	An acceptable characterization and cleaning of the dataset that summarizes all details from source to fields. An adequate accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used.	Provides a vague or incomplete explanation of the purpose of hyperparameter tuning, indicating a need for improvement. Provides a vague or incomplete explanation of hyperparameter tuning techniques, indicating a need for improvement.	An acceptable interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results state that the models are adequate and require more depth. An adequate justification is provided.
Fail (<39%)	An impecunious introduction to problem description that fails entirely to motivate the problem. An impecunious specification of objectives.	An impecunious characterization and cleaning of the dataset. An impecunious obtained based on the training and testing of ML models using three logical splits. Cross-validation is not used.	No meaningful response or completely incorrect information. No meaningful response or completely incorrect information.	Poor interpretation of results, missing code comments, unsatisfactory conclusions, and incorrect citations.

The Irish Grading System

The grading system in CCT is the QQI percentage grading system and is in common use in higher education institutions in Ireland. The pass mark and thresholds for different grade bands may be different from what you have experienced in the higher education system in other countries. CCT grades must be considered in the context of the grading system in Irish higher education and not assumed to represent the same standard the percentage grade reflects when awarded in an international context.

Please review the CCT Grade Descriptor available on the module Moodle page for a detailed description of the standard of work required for each grade band, and review the marking criteria outlined in this assignment brief for a breakdown of the marking criteria for this specific assignment.

Additional Information

- Lecturers are not required to review draft assessment submissions. This may be offered at the lecturer's discretion.
- In accordance with CCT policy, feedback to learners may be provided in written, audio or video format and can be provided as individual learner feedback, small group feedback or whole class feedback.
- Results and feedback will only be issued when assessments have been marked and moderated / reviewed by a second examiner.
- Additional feedback may be provided as individual, small group or whole class feedback. Lecturers are not obliged to respond to email requests for additional feedback where this is not the specified process or to respond to further requests for feedback following the additional feedback.
- Following receipt of feedback, where a student believes there has been an error in the marks or feedback received, they should avail of the recheck and review process and should not attempt to get a revised mark / feedback by directly approaching the lecturer. Lecturers are not authorised to amend published marks outside of the recheck and review process or the Board of Examiners process.
- Students are advised that disagreement with an academic judgement is not grounds for review.
- For additional support with academic writing and referencing students are advised to contact the CCT Library Service.
- For additional support with subject matter content students are advised to contact the CCT Student Mentoring Academy
- For additional support with IT subject content, students are advised to access the CCT Support Hub.